



**Supplementary Figure S9.** The flow diagram illustrates the hybrid feature selection process. The first step of the hybrid feature selection involved getting rid of highly correlated features and features with more than 80% missing values were also eliminated. Next, an initial round of feature selection was performed using the feature selection by model method with four different machine learning models : L1-based linear regression, Extremely Randomized Trees (Extra-trees model), Random Forest, and Gradient Boosting classifiers. A second round of feature selection is then carried out using the wrapper-based feature selection method (Recursive Feature Elimination (RFE) with cross validation) on the pre-selected set of features. Finally, the Union selector and a base gradient boosting classifier resulted in the best subset of top 25 features to be used for subsequent analysis.